

Multimodal fusion of IMUs and EPS body worn sensors for scratch recognition

Zygimantas Jocys
University of Sussex
Falmer, UK
z.jocys@sussex.ac.uk

Arash PourYazdan
University of Sussex
Falmer, UK
A.Pouryazdan@sussex.ac.uk

Daniel Roggen
University of Sussex
Falmer, UK
d.roggen@sussex.ac.uk

ABSTRACT

Itching may be caused by different skin diseases. In order to develop and evaluate how much itching affects a person's daily life, it is useful to develop automated means to recognize the action of scratching. We present an investigation of sensors and algorithms to realize a wearable scratch detection device. We collected a dataset, where each user wore 4 IMU (Inertial Measurement Units) sensors and one EPS (electric potential sensor). Data was collected from 9 users, where each user followed a 40 minute protocol, which involved scratching different parts of head, shoulder and leg, as well as other activities such as walking, drinking water, brushing teeth and sitting next to the computer. The dataset contains 813 scratching instances and 5h 15 min of recorded data. We investigate trade-offs between number of devices worn, and hence comfort, and recognition performance. We trained k-NN and Random Forest using between 1 to 5 of the best features per channel to detect scratches. We conclude that scratch can be detected with 80.7% by using Random Forest on hand coordinates, which require 4 devices. However, f1 score of 70% can be achieved with k-NN with IMU and EPS data, which only requires 1 device. Moreover, fusion of IMU data with EPS data improved the accuracy and reduced the standard deviation between the folds. This expands the state of the art by opening up new trade-offs between accuracy and comfort for future research in skin conditions.

KEYWORDS

Scratch detection, feature selection, machine learning, activity recognition, sensor technologies.

ACM Reference Format:

Zygimantas Jocys, Arash PourYazdan, and Daniel Roggen. 2020. Multimodal fusion of IMUs and EPS body worn sensors for scratch recognition. In *Atlanta '20: Pervasive Health 2020, May 18–20, 2020, Atlanta*, 11 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUCTION

Itch is a condition which affects a substantial group of people. This condition may be caused by scabies, atopic dermatitis, kidney failure, it can also be a symptom of a malignant condition, such as

lymphoma [18]. Scratching occurs as a result of itching, but also increases itch while scratching. Scratching also produces wounds on the skin, which create even more discomfort. Itch and scratch are correlated [7].

Furthermore, atopic dermatitis causes high intensity of itch followed by a period of respite. The amount of scratching which occurs over night and what impact it has on the skin cannot be evaluated by the user himself as he is not consciously aware of it. We need a system which could, during such flares, warn the user to apply the treatment on the affected area, to reduce the sensation of itch. It would improve the sleep quality by reducing the intensity of the cycle: itch, scratch, more intense itch.

Moreover, automatic monitoring for atopic dermatitis clinical trials would enable more accurate measurement of how the frequency and intensity of scratching changes before and after the user starts treatment.

In previous research, acceleration sensors with simple models such as logistic regression [13] and more **complex Deep Learning models** [10] showed good success in detecting scratch. In this research we present a new systematic approach towards the evaluation of needed sensor modalities and the complexity of machine learning models to successfully detect scratches. The novelty of this work is:

- Survey of previous scratch detection work to evaluate - what dataset is required to represent the daily lives of humans.
- A new dataset using 4 Inertial Measurement Units, 1 Electric Potential Sensor and hand coordinates with 10 activities and 9 users with 813 scratches and a total of 5h15 min of recorded data. The novelty of this dataset lies in more sensor modalities, to push the machine learning models to the limits and evaluate how well slightly different scratches can be differentiated from other activities. The new dataset has more scratch locations as well as a more realistic NULL class.
- We observe that EPS data combined with IMU data reduces the error variance between the folds and increases the accuracy of human activity recognition.
- Comparison of k-Nearest Neighbour and Random Forest by using different number of features scored by Mutual Information. Previous work has not explored feature selection for scratch detection from different modalities.
- Exploration of trade-offs between accuracy, comfort and number of devices.

2 RELATED WORK

In previous work, itch [2] was described as irritating feeling that causes the desire to scratch. Chronic itch can be a reaction of skin diseases, kidney failure, cancers and neurological disorders. Given

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Atlanta '20, May 18–20, 2020, Atlanta

© 2020 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

that the disease might need urgent medical attention, scratches should be monitored as it could lead to a faster disease identification.

In the early days, scratch was observed manually, without any technological help. In the research [8], scratching was monitored on 40 patients in the hospital by the staff. During this research project the amount of scratching occurring before and after the treatment was noted. Observing the change of scratch quantity before and after the treatment allows to evaluate how much the treatment reduces the occurrences of scratches. That system is not suitable to be deployed on a large group of people (hundreds).

Another approach is to video record nocturnal scratches using an infrared camera. In one experiment, which was done with 7 users [6], the infrared camera was used, so that the user would not be interrupted during the night and the scratches are annotated later.

Currently, the most common approach to classify activities is to use accelerometers [3]. It allows for the monitoring of activities without invading user privacy. Not only does it allow for systems to be deployed on many users simultaneously, but accelerometers also enable real-time and real-world monitoring without the need of an observer to invade the subjects privacy. Triaxial accelerometers [13] on the wrist for scratch deployment were noted successful and achieved high sensitivity (0.96) and specificity (0.92). In that research 12 people participated where they were monitored in a controlled environment. The duration of the dataset collection was 140s per user. Each person scratched the back of their head, leg, elbow and later walked and rolled around. **A similar work was done by detecting scratch using mobile systems [9], where they achieved 90% accuracy.**

However, even though a single modality has shown good results, there are many new sensors, which could potentially be incorporated in such a system and increase the accuracy of the scratch detection.

Acoustic sensors [11] have been used for scratch detection. The sound sensor uses body conduction to record the sound from the wrist. The dataset was collected on 4 volunteers during the night and was collected over 6h. As the room is silent, it is easier to distinguish between scratch or no scratch, where they achieve a 0.98 "determination ratio".

Electric potential sensing (EPS) was reported as a new modality for scratch sensing [14]. This work only reported a visual observation that EPS signals are correlated to scratch, but the authors did not demonstrate any automated activity recognition based on this modality.

Machine learning models vary in complexity, computation time and their ability to generalise. Logistic regressions [13] and Recurrent Neural Networks [10] have been used. In the research [10], 24 atopic dermatitis patients spent 2-5 nights in the hospital and wore 2 accelerometer devices (GeneActiv, Activinsights Ltd.) on each wrist. Only two classes were labeled: scratch or null class, and as the data needs to be split for each class equally a lot of recording data was lost for the training of the model. There are still room for new machine learning models and much progress to be done with the current techniques.

In the previous work all the currently collected datasets for scratch detection are either very short (150s) or very simplistic (2 classes: scratch or null class), which does not allow to evaluate, how

the system would perform in a more realistic scenario. Moreover, given that in each research there is one sensor modality per dataset, there is a need to explore different fusion options between different sensors.

We incorporate 4 IMUs, extract hand coordinates and use EPS for scratch detection during the dataset collection. Moreover, there is a need of a challenging dataset which has multiple classes and represents a more realistic world. A more complex dataset would allow to evaluate how different sensor modalities contribute towards an efficient and effective human activity recognition.

3 MULTIMODAL SCRATCH DATASET

A new dataset was collected from 9 users. It contains 2 classes and 10 subclasses, which are shown in the Table 1. Each user wore 4 IMU sensors and 1 EPS sensor.

Class	Subclass
Scratch	Top of the head
	Back of the head
	Side of the head
	Shoulder
	Leg
Null	Null
	Drink water
	Brush teeth
	Wash hands
	Walk

Table 1: 10 classes of the dataset

As it can be seen from Table 2 in total we recorded 40 min of scratching and 5h 15min of data. The dataset has 10 classes and some activities have been done simultaneously in order to find the limits of the machine learning models and it's ability to learn subtle difference between very similar activities. In our case the subtle differences are scratch top of the head, back of the head and side of the head.

Dataset parameters	
Age	22-40
Males	6
Females	3
Scratching instance	813
Total scratching time	40min
Collection time per user	35min
Total dataset time	5h15min

Table 2: Parameters of the dataset

Participants had different hair lengths and there were 6 males and 3 females. The diversity of users' biological features are needed to evaluate if the model is able to generalise enough to detect the same activities on unseen users. There were no overweight or underweight participants, all the participants had average body shapes. It is also important to have a long duration and a diverse null class, so that the model is able to learn the difference to distinguish scratch and other activities.

3.1 Sensor set up

The IMU is based on BlueSense technology [15]. BlueSense gives raw acceleration, rate of turn and magnetic field, and can also provide the quaternion data, which encodes the orientation of the device. Moreover, the quaternion data from the 4 IMUs can be used to get the approximate hand coordinates (explained in the section 3.1.2). Furthermore, BlueSense can be extended with expansion boards comprising additional sensors. In this work, we used an expansion board with an EPS sensor [14].

3.1.1 EPS and IMU. The sensor configuration for each user is shown in Figure 1 (c). This configuration was chosen in order to get hand coordinates in Cartesian system, the charge of the electric field and the IMU data.

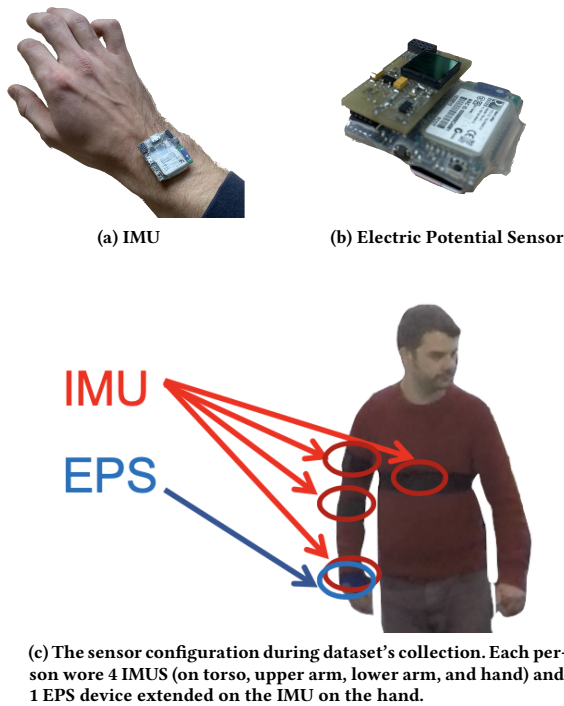


Figure 1: Imu(a) and EPS(b) are used for the data collection. The sensors were attached to the body using 4 straps.

In this work, a device is a single object, IMU with EPS is one device, sensor is a sensor modality (IMU and EPS is two sensors). We refer to a *channel* as one output of a sensor. Sensors can have one or more outputs. For example, in this work an IMU has 13 channels. They are: 3 channels of acceleration along the x, y, z axis; 3 channels of rate of turn along the x, y, z axis; 3 channels of magnetic field along the x, y, z axis; and 4 channels indicating the device orientation in quaternions. EPS has only one output channel which represents the electric potential.

Electric Potential Sensor, shown in Figure 1 (b), streams the voltage data using ADC channel to Bluesense [15]. Once the electric field is disrupted [14] it can be observed in the signal. Also, EPS

sensor has the ability to detect 50 Hz grid voltage when the computer is connecting to the power source, which allows to accurately detect typing to a computer, once it is connected to a power source.

3.1.2 Hand coordinates. Hand coordinates were computed after the dataset collection. The quaternion data captures the orientation of each sensor. The sensor positioning is seen in Figure 1 (c). By getting the orientation of each sensor and by using a vector for each limb to calculate the coordinate of each joint, all the Cartesian coordinates of each joint are summed to get the hand coordinate.

3.2 Dataset collection

This work was approved by University of Sussex Ethical Committee application N. ER/ZJ70/1. The participants were recruited in the Engineering and Informatics building for the participation in the data collection. During the dataset collection participants needed to follow the protocol.

We defined a protocol, displayed in Figure 3, which allows to collect a wide variety of scratches and is embedded in a naturalistic scenario comprising other activities forming a null class. In particular, we collected here a dataset comprising three different scratch locations on the head (top, side, back), and scratch locations on other parts of the body (leg, shoulder). These scratches were performed with 2 different intensities (intense and moderate intensities). In order to include a realistic null class in the dataset, we asked participants to walk about in the 20s in the office in between each scratch. In addition, once all the scratches were performed we asked participants to perform a number of other hand gestures, including simulating washing hands and brushing teeth, drinking, writing at a computer, as a way of including a more realistic set of activities to evaluate how well scratch can be distinguished from other activities of daily living.

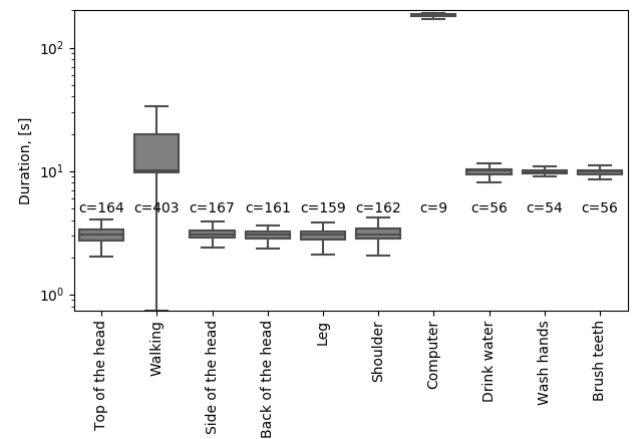


Figure 2: The duration of each activity in the collected dataset. The value c denotes the number of occurrence of each activity across the whole dataset.

The dataset collection is a tedious process and has many activities. An application was used to show the current and the next activity

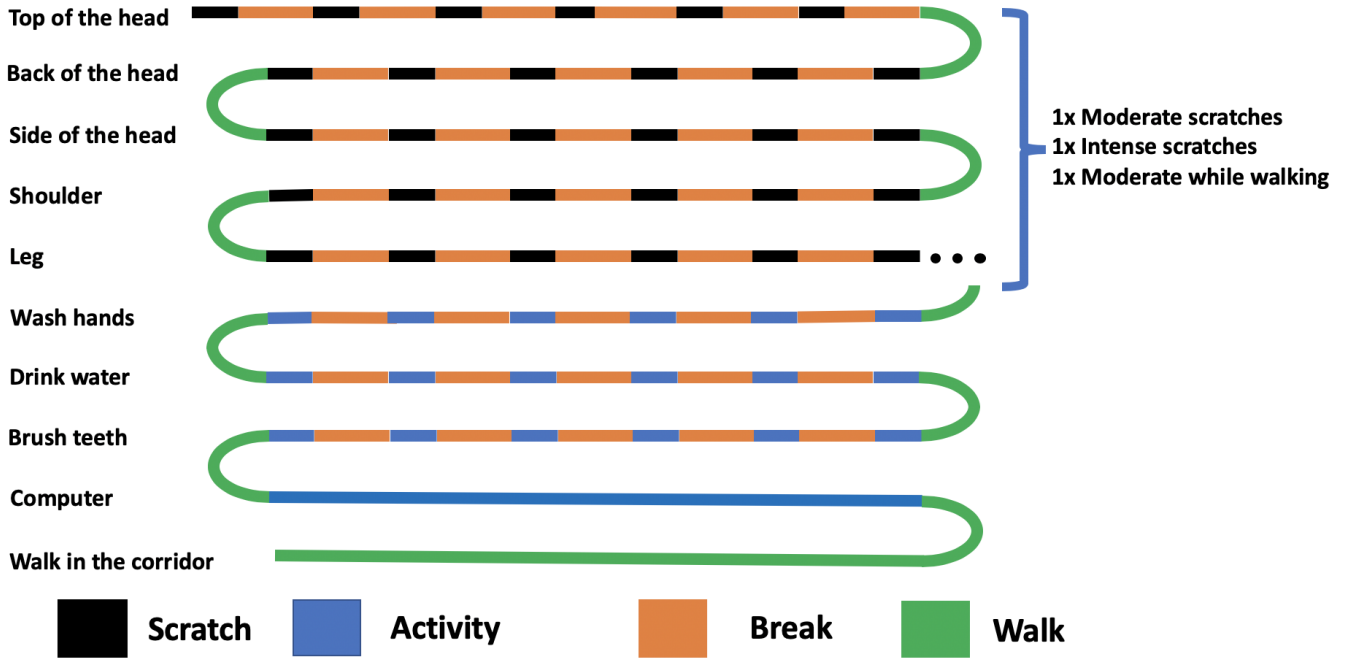


Figure 3: Protocol of the dataset collection. First, the scratches are done at three different intensities and all the other activities for the null class are done after wards. The scratch occurs for 3s and then there is a 10s break. It is done 6 times and before the next activity the 20s walk occurs.

on an iPad, using the application "Seconds" [1]. Undesired deviation from the given protocol occurred during the collection. However, the labels of activities were adjusted after the data collection to adjust to the deviations.

4 HUMAN ACTIVITY RECOGNITION PIPELINE

Human activity recognition process is a process that requires a specialised pipeline for each case. The pipeline we used is shown in Figure 4. This pipeline is used for the human activity recognition.

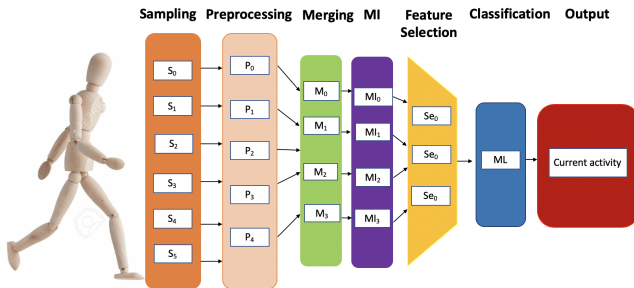


Figure 4: Human activity recognition pipeline for scratch detection. The data is sampled from the sensors, which is then preprocessed. And the features, which have the highest MI score are used for the classification task.

4.1 Data Cleaning and Pre-processing

The data was collected on 5 devices (4 IMUs and 1 EPS). However, with an extra development of the firmware, the IMU and EPS could be deployed using only one device. Thus, in the later stages the number of devices needed to have EPS and IMU is 1.

The sampling rates of the IMU and EPS are 100Hz and 1kHz. After sampling, the data is stored on the local SD cards. All the devices were synchronised, however they did not log the data at the exactly same time. Due to the different time stamps some interpolation technique should be considered. As the quaternion data has a value of an angle, linear interpolation is not possible. Thus, an ASOF function was used. ASOF merges the data on nearest timestamp rather than equal timestamp. A time delta equal to 10ms, was chosen: if the nearest timestamp is further than 10ms then it does not choose the nearest value and assigns NaN.

The EPS is very sensitive and the collected data has a great amount of fluctuation. Thus, a low pass Butterworth filter is applied to smooth the signal and then it was resampled to match the 100 Hz IMU's frequency.

4.2 Channels

Additional data, apart from the sampled data is computed, to have more information:

- The hand coordinates, which are described in section 3.1.2.
- In the formula, m is any modality that has the x, y, z projections and m_{xyz} is the magnitude, which is calculated to see if the magnitude can enable to achieve a higher performance.

$$m_{xyz} = \sqrt{m_x^2 + m_y^2 + m_z^2} \quad (1)$$

The magnitude is computed using formula 1 was computed for acceleration, rotation and hand coordinates.

Acc _x	Quat ₀	Gyr _x	Hand _x	EPS
Acc _y	Quat ₁	Gyr _y	Hand _y	-
Acc _z	Quat ₂	Gyr _z	Hand _z	-
Acc _{xyz}	Quat ₃	Gyr _{xyz}	Hand _{xyz}	-

Table 3: Channels. All the channels that are used in activity recognition are displayed including acceleration, rotation, orientation, hand coordinates and the EPS.

Afterwards, sliding windows of the timeseries are generated with window length of 0.4s. This time was chosen, due to the fact that scratch is an activity which occurs for a short amount of time.

4.3 Feature Extraction

From the features shown in Table 3, we need to extract features from the sliding windows.

The features where chosen for this case are show in Table 4. Mean and variance enables to define the distribution of the series. The percentiles allows to detect the key points in the distribution and avoids the outliers, contradictory to the minimum and maximum functions. The mean crossing rate and zero crossing rate are used to evaluate how periodic the channel is.

Domain	Features
Statistical	Mean
	Variance
	Percentiles 25%
	Percentiles 50%
	Percentiles 75%
	Percentiles 90%
	Mean crossing rate
Frequency	Zero crossing rate
	Energy

Table 4: List of features

4.4 Feature Selection

Two of the most common ways to select the most important features are filter and wrapper methods. As we have 17 channels, there are 170 unique features. Wrapper method would take unreasonable amount of time to find the best combination of features. Thus, filter method, that uses Mutual Information [5] algorithm, is used to select the features carrying the highest amount of information.

$$I(X; Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p_{(X,Y)}(x, y) \log \left(\frac{p_{(X,Y)}(x, y)}{p_X(x) p_Y(y)} \right) \quad (2)$$

where $p(X, Y)$ is the joint probability mass function of X and Y , and p_X and p_Y are the marginal probability mass functions of

X and Y respectively, where X is the bin of the feature and Y is the class. For each feature we created 100 equal size bins and did a small correction by adding 10^{-10} to the division so that when there are empty bins, division by zero is possible.

4.5 Machine Learning

As the range of the data varies highly, all the extracted features are normalised as shown in equation 3.

$$n_{normalized} = \frac{n_{feature} - \mu}{\sigma} \quad (3)$$

In the equation μ is the mean of the feature in the training set, n_{raw} is the feature and σ is the standard deviation of the feature in the training set.

4.5.1 k-Nearest Neighbour. k-Nearest Neighbour is a model which is fast to train and has a proven record of successful applications in many areas. On the other hand, it is slow to compute predictions and is very susceptible to outliers. Thus, choosing too many correlated features or features that do not bring valuable information towards decision making, would only diminish the model's ability to recognise activities.. The value of k was chosen to be 100 and the sklearn implementation was used.

4.5.2 Random Forest. Random Forest is ensemble based learning method. This model has seen much success in regression and classification tasks. Given that the decision trees are able to distinguish the important features, it means that a bigger number of features will lead to more accurate predictions. For Random Forest it was chosen to use 100 trees. In this case we use the Random Forest algorithm from sklearn Python library.

4.6 Measuring performance

The goal of this project is to have a universal system for all the users. To achieve this goal, we are using 3-fold cross-validation. During the cross-validation the users are grouped into 3 groups of 3 users, where during each validation it is tested on the groups of 3.

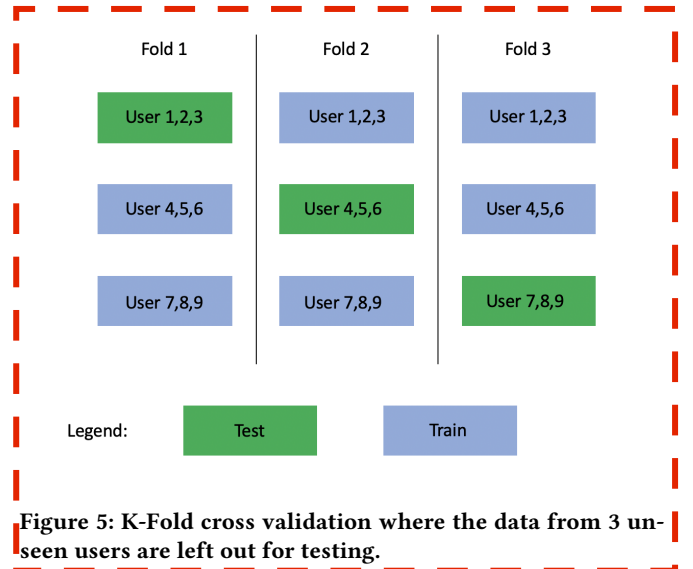


Figure 5: K-Fold cross validation where the data from 3 un-seen users are left out for testing.

ML model	Modalities	Number of devices	Number of features				
			1	2	3	4	5
k-NN	xyz	4	38.01±0.7%	59.3±2.5%	61.2±1.9%	62.7±1.9%	62.1±2%
	IMU	1	41.4±6.7%	45.4±8.3%	46.3±7.8%	46.4±8%	47±7.3%
	IMU+EPS	1	43.2±2.1%	50.3±2.4%	52±2.1%	52.7±2.2%	53.4±2.4%
	IMU+xyz	4	43.8±6.2%	52.5±0.9%	55.9±3.2%	55.9±2.8%	56.3±28%
	IMU+xyz+EPS	4	47.9±1%	52.2±1.7%	54±2%	54.7±0.2%	55.6±2%
Random Forest	xyz	1	55.1±2.5%	66.6±3.5%	66.7±3.2%	66.8±3.2%	66.9±3.2%
	IMU	1	45.6±8.2%	47.7±8.4%	48.1±8.5%	48.4±8.6%	49±8.5%
	IMU+EPS	1	45.2±4.2%	49.6±5.4%	50.6±5.7%	51.8±6%	52.1±5.8%
	IMU+xyz	4	50.6±3.3%	52.6±4%	51.96±4.5%	52±4.6%	53.8±4.6%
	IMU+xyz+EPS	4	56.3±2.2%	57.4±3%	57.8±3.1%	57.8±3.1%	57.8±3%

Table 5: Result table of the 10 class classification with the displayed Macro f1 score. It can be noted that the better performance overall is achieved with the Random Forest model. The xyz is a simplified notation for the hand coordinate data.

In this research project, confusion matrix and Macro f1 score are chosen to evaluate the performance of the model and its ability to generalise. It was chosen to use the Macro score in order to see how well it recognise each class.

$$F_{1c} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

$$F_{1Macro} = \sum_{c=1}^n F_{1c} \quad (5)$$

The Macro f1 score is calculated using equation 5. The score for each class affects the overall results. Due to the class imbalance in the dataset (40 min from 5h15 minutes are scratches and the rest is null), using the Micro f1 score we would achieve a high performance if it gets the majority of null class.

4.6.1 *Fine activity recognition.* First, the models will be tested to see how well they can classify the 10 classes. As there are very similar classes and there multiple activities happening at the same time this dataset was created to push the limits of machine learning models.

4.6.2 *Coarse scratch / non-scratch activity recognition.* The 10 class classification task is meant to push the machine learning model to distinguish the subtle difference between the activities. By evaluating how well scratch is distinguishable from a complex null class, we will be able to see how well scratch could be detected in realistic scenarios. For this part of the experiment, the model will not be retrained, but the labels will be changed to either scratch or null. This means that if "scratch top of the head" is confused with "scratch back of the head" it will not reduce the f1 score. The constructed classes of scratch and null class are shown in table 1.

5 RESULTS

5.1 Most important features

Using Mutual Information a heatmap of most important features was produced. It is shown in Figure 6. For each channel 5 features (which carry the highest amount of information) have been selected and are displayed in table 6. As it can be seen, variance carries the biggest amount of information for most channels and the

percentiles show good performance as well. Looking at the results in the Figure 6, EPS carries the smallest amount of information as a channel. Mean crossing rate and zero crossing rate are not informative features in most cases.

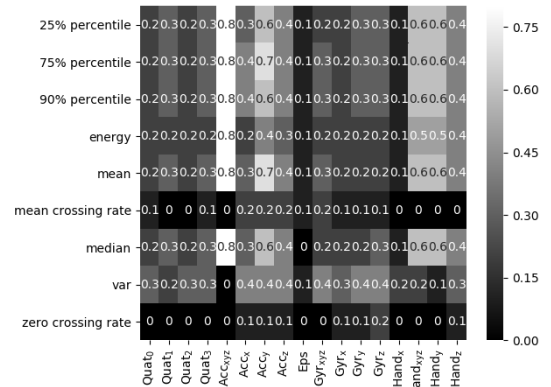
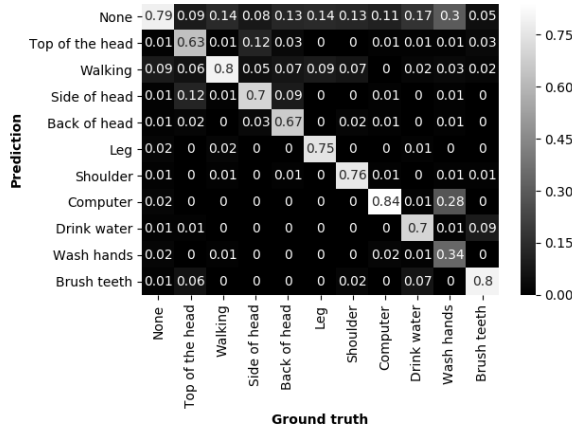
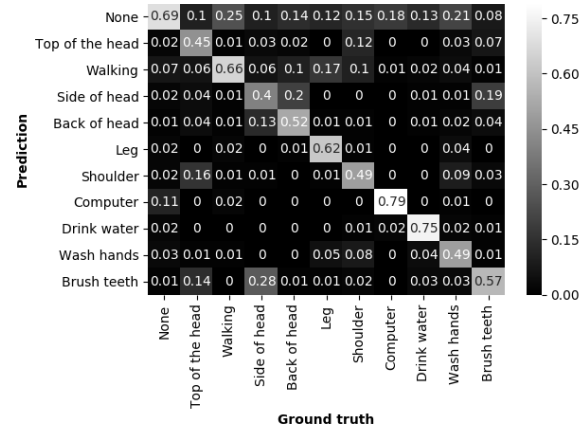


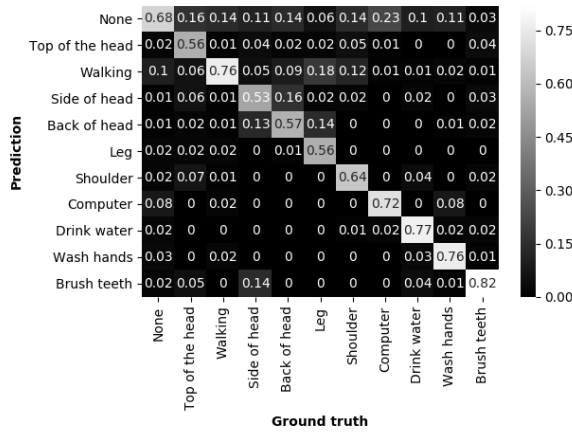
Figure 6: Feature heat map based on Mutual Information. The lighter shade indicates that the feature of the channel carries more information than the darker shade.



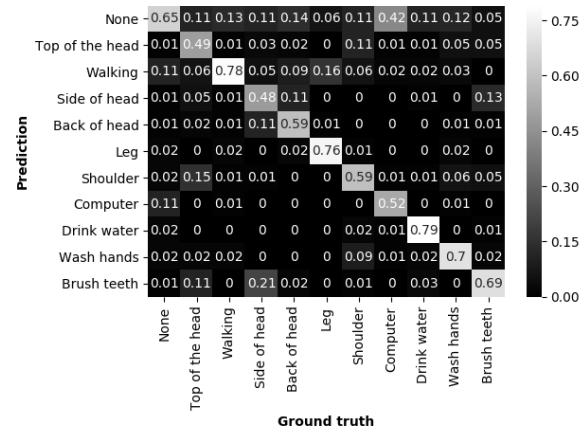
(a) hand coordinates



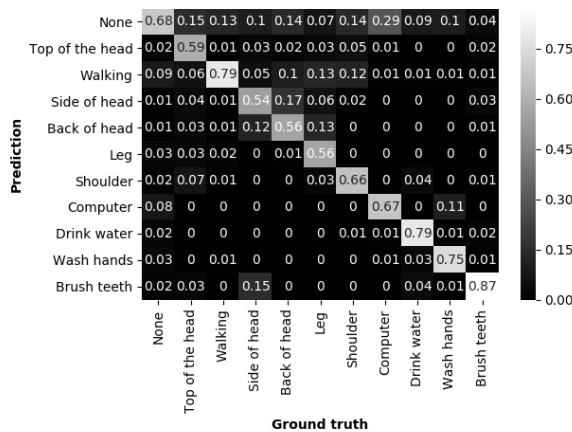
(b) IMU channel



(c) IMU and hand coordinates



(d) IMU and EPS channel



(e) Hand coordinates, IMU and EPS

Figure 7: Confusion matrices when inputs are 2 features per channel and the model is Random Forest.

Channel	f1	f2	f3	f4	f5
Acc _x	var	90 %	75%	mean	median
Acc _y	mean	75 %	90%	median	25%
Acc _z	25%	mean	median	var	25%
Gyr _x	var	25 %	90%	75%	median
Gyr _y	var	90 %	25%	75%	median
Gyr _z	var	90 %	75%	25%	median
Quat ₀	var	90 %	75%	25%	median
Quat ₁	75%	90 %	median	mean	25%
Quat ₂	90	75%	median	25%	mean
Quat ₃	var	90 %	75%	median	mean
Acc _{xyz}	mean	median	energy	25%	75 % %
Gyr _{xyz}	var	90 %	75%	mean	median
Hand _x	var	90 %	25%	75%	median
Hand _y	25%	median	75%	90 %	mean
Hand _z	25%	mean	median	75%	90 %
Hand _{xyz}	75 %	90 %	median	25 %	mean
Eps	mx0	var	90 %	mean	75 %

Table 6: Selected features for each channel

5.2 Fine activity recognition results

In the Table 5, the results are presented for 10 class classification task using k-NN and Random Forest while trained on a different number of features from 1 to 5 per channel. The accuracy is shown as a mean and standard deviation between the 3 folds.

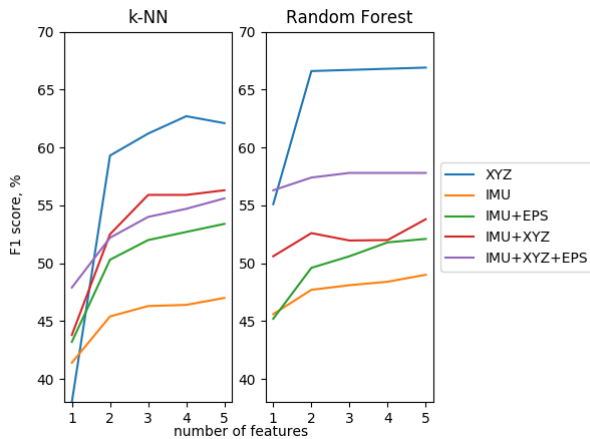


Figure 8: f1 score vs number of features for 10 class classification

5.2.1 k-NN. The best results have been observed when solely hand coordinates with 4 features per channel are used. The f1 score reached 62.7%. To achieve this result, 4 devices are required. On the other hand, with solely IMU and EPS, 52.3% can be achieved, where 1 device is required.

Combining hand coordinate with IMU or IMU+EPS data only worsens the results.

A result of 53.4% f1 score was achieved with 1 device (IMU and EPS), while with IMU solely, the best achieved result was 47%.

5.2.2 Random Forest. Random Forest showed better performance overall. The best performance was seen when using hand coordinates with 4 features per channel where an f1 score of 66.9% was achieved for 10 class classification. However, this task requires 4 devices on the body. It can be observed that performance was growing with the amount of features introduced.

A result of 52.9% f1 score was achieved with 1 device (IMU and EPS). Nonetheless, with IMU solely, the best achieved result was 49%.

In the Figure 7, the confusion matrix shows which classes are mostly confused when it is tested on RF model using 2 features per channel. In all the confusion matrices it can be noted that "Scratch the top of the head", "Side of the head" and "Back of the head" are greatly confused. In Figure (b) it can be seen "Brushing teeth" is confused with "Side of the head: and vice versa, and in Figure (a) "Washing hands" are confused with "Typing to a computer". When both modalities are combined as seen in Figure (c), the performance on those two classes drastically increase. When EPS is introduced, in Figure (d) it can be seen that the amount of errors decrease in the scratch classes, compared to when solely IMU is used.

5.3 Coarse scratch / non-scratch activity recognition results

It is very challenging to classify 10 classes. However, to evaluate how well scratch is recognisable, it is enough to distinguish scratch from any other activity. In the Table 7, the results are presented for the binary classification task.

5.3.1 k-NN. The best result was achieved again by using the hand coordinate data. It achieved a f1 score of 77.6% and with a single device (IMU and EPS) an f1 score of 70% with a standard deviation of 2.9%. By solely using the IMU a f1 score of 62% was reached with a standard deviation of 8.6%

The EPS allows to reduce the standard deviation between the folds compared to the IMU results. However, 4 devices are needed with extracted hand coordinates to achieve the highest results.

5.3.2 Random Forest. The best score was achieved by using Random Forest with 5 features per channel. It reached a f1 score of 80.7% for scratch detection. With just IMU and EPS an accuracy of 69.8% can achieve and the standard deviation between the folds was only 7.2%. Solely IMU reached 63.8% with a standard deviation of 10%.

The same behavior is observed as with the k-NN. The EPS allows to reduce the standard deviation between the folds compared to the IMU results. However, 4 devices are needed with extracted hand coordinates to achieve the highest results.

ML model	Modalities	Number of devices	Number of features				
			1	2	3	4	5
k-NN	xyz	4	55.89±2.7%	75.7±3%	77.2±2.5%	77.6±2.7%	77.5±2.4%
	IMU	1	59.3±5.5%	60.8±8.5%	61.8±8.4%	61.8±8.4%	62±8.6%
	IMU+EPS	1	63.1±2.7%	67.8±2.9%	69.3±2.9%	69.6±2.8%	70±2.9%
	IMU+xyz	4	74.9±1.4%	73.5±78.3%	72±3.6%	71.8±3.8%	70.6±5.11%
	IMU+xyz+EPS	4	66±4.4%	70.3±4.5%	72.6±4.5%	72.7±4.5%	72.9±4.7%
Random Forest	xyz	4	67.7±2%	80.1±2.7%	80.4±2.3%	80.7±2.7%	80.7±2.6%
	IMU	1	62.3±6.3%	63.3±8.7%	63.3±9.2%	63.8±10%	63.8 ±10%
	IMU+EPS	1	62.09±1.7%	68.2±6.4%	69.5±6.8%	69.5±7.6%	69.8±7.2%
	IMU+xyz	4	66.2±3.9%	69.6±2.7%	69.6±2.5%	67.9±3.3%	69±3.3%
	IMU+xyz+EPS	4	71.6±3.3%	73.5±3.4%	73.7±3.3%	73.5±3.1%	73.1±3.6%

Table 7: Result Table of binary classification with the displayed Macro f1 score. It can be noted that the better performance overall is achieved with the Random Forest model. The xyz is a simplified notation for the hand coordinate data.

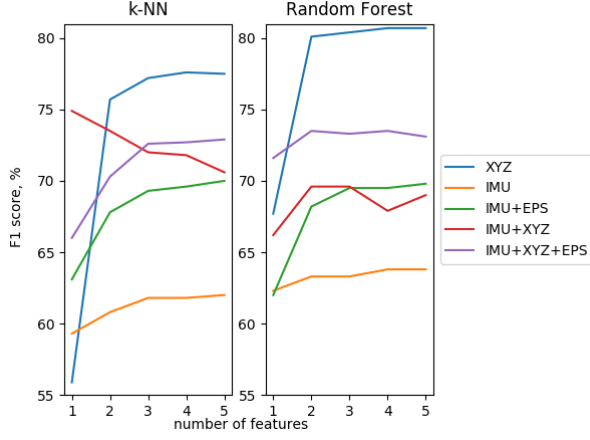


Figure 9: f1 score vs number of features for binary classification

6 DISCUSSION

6.1 Baseline Results

Currently, the standard approach towards human activity recognition is predicting activity from IMU data. In this work, the baseline results are achieved by using the extracted features from the IMU data. Moreover, these results only require 1 device to be deployed.

The baseline results of the new dataset are:

- For 10 class classification with k-NN the best result is $47 \pm 7.3\%$.
- For 10 class classification with RF the best result is $49 \pm 8.5\%$.
- For binary classification with k-NN the best result is $62 \pm 8.7\%$.
- For binary classification with RF the best result is $63.8 \pm 10\%$.

6.2 Multimodal Fusion

The fusion of data between IMU, EPS sensors and the extracted hand coordinates does not always result in better performance.

6.2.1 IMU and EPS. IMU and EPS requires 1 device on the wrist. Fusing the data and classification from IMU and EPS data provides a slightly better performance than solely using IMU with Random

Forest and k-NN. Using the binary classification task with IMU and EPS it resulted in 70% using k-NN and 69.8% using RF. Moreover, with k-NN it reduces the standard deviation between the folds from 8.6% to 2.9% and from 10% to 7.2% using Random Forest. Compared to the baseline results there is a small improvement in the f1 score. In both cases, fusing the IMU and EPS data reduces standard deviation and improves the results using both models, but not significantly. The reduction in standard deviation of f1 score between the folds, shows that the model's ability to recognise the activities is good for a diverse group people. Moreover, as the EPS can be deployed on the IMU located on the hand, it does not create more discomfort.

6.2.2 IMU and hand coordinates. For 10 class and binary classification combining these modalities give better results than using solely IMU, but worse than using solely hand coordinates. Using RF and binary classification IMU achieved $63.8 \pm 10\%$, just with hand coordinates the result is $80.7 \pm 2.6\%$ and when combined it achieved $69.6 \pm 2.5\%$. The same behavior can be observed with k-NN and the 10 class classification task.

Each activity is associated more with certain location relative to the torso, rather than specific movements. Thus, additional information could be redundant and cause the accuracy to decrease. In this case, doing the hand coordinate fusion with IMU does not provide with any gains, from the results achieved solely from the hand coordinates.

6.2.3 IMU, hand coordinates and EPS. Combining all the modalities is outperformed solely by using only hand coordinates. For binary classification task using RF and binary classification IMU achieved $63.8 \pm 10\%$, just with hand coordinates the result is $80.7 \pm 2.6\%$, IMU with hand coordinates it achieved $69.6 \pm 2.5\%$ and with the IMU, hand coordinates and EPS the model was able to achieve a $73 \pm 2.6\%$ f1 score.

The dataset is biased to the location. It means that each activity in the dataset is associated with a certain location relative to the torso and this particular feature works exceptionally well with this dataset. The fusion of the dataset does not bring a drastic improvement in the accuracy, but reduces significantly the standard

deviation between the folds, compared to the baseline results. However, it still does not outperform the achieved results when solely hand coordinates are used.

6.3 Number of features

As it can be seen from Figures 8 and 9, there is an increase in the accuracy, but afterwards the results plateau. It might be caused by the mutual information algorithm which does not take into account that certain features are correlated and that the redundant information does not improve the performance. This was observed for both classifiers, except for the exception of k-NN with IMU and hand coordinates.

6.4 Comfort and accuracy trade-off

To deploy a scratch detection system for clinical trials the smallest number and the least invasive device must be chosen. Sensor set up with 4 devices (shown in Figure 1) could not be used in any medical or clinical study. This set up would make daily activities uncomfortable and there would be a higher risk of failure due to the high number of devices in use.

For this data collection 4 devices were used. In general more information gives better results. In the Figure 5 and 7 it is shown that increasing the number of devices produced better results with the exception of k-NN and hand coordinate. However, discomfort is a major drawback for a deployment of a human activity recognition system. If comfort is the priority: 1 device with IMU and EPS on the wrist might be enough with a 70 % accuracy, if accuracy is the priority then with 4 devices (as shown in Figure 1) an 80.7% score could be achieved solely from hand coordinates.

6.5 Future Work

In section 5, it can be noted that hand coordinates relative to the torso are needed to achieve the best performance for this dataset. To have a comfortable system new localisation techniques should be explored, so that the sensors could fit on one wrist. We suggest to explore localisation techniques, such as ultra wide band. As an example PosXYZ [4] needs only 2 devices (slave and the master) and could be deployed in such a system. The slave device would be on the wrist and master device would be the reference point. Using Ultra-wide-band technology, the location of the wrist compared to the torso would be computed and could be used for human activity recognition. That way the hand coordinates would enable to achieve higher accuracy, without needing four straps on the torso, upper arm, lower arm and wrist.

Exploring different feature selection techniques, such as MRMR [17], could lead to higher performance. MRMR is a minimum redundancy feature selection algorithm, which also takes into consideration how redundant the feature is compared to other selected features. Moreover, to see what accuracy can be achieved for this dataset, investigating Deep Learning models such as DeepConvLSTM [12], could produce substantially higher accuracy results.

Finally, new hardware could be developed for head scratch detection, such as EPS based glasses, which could work as proximity sensor to detect when the hand is nearby, thus increasing the recognition of head scratches.

A feasibility study was done to evaluate the detection of smoking in daily life [16]. This work gives the insight that the data collection should have incorporated active learning to confirm if the activity is happening during the data collection. The 4 participants were smokers and they needed to tap the sensor to flag when the smoking happened. However, flagging the activity requires being conscious of doing the activity.

7 CONCLUSIONS

During this experiment we explored how the fusion of different sensor modalities are contributing towards accurate scratch detection using different number of features per channel and common machine learning models, such as k-NN and Random Forest. For this task, a dataset was collected with 10 different activities to investigate the limitations of each model and explore this trade-off between the number of sensor modalities, number features and machine learning models.

The key results are:

- Best baseline result for detecting scratch with a simple IMU reached 63.8% f1 score using RF, which needs only 1 device.
- Best result using one device was 70% f1 score for scratch detection. It was achieved by using k-NN with IMU and EPS data.
- Best result overall was 80.7% f1 score. It was achieved for the binary scratch detection from hand coordinates using RF model, which requires 4 devices.
- Fusing EPS data with IMU data consistently increased the accuracy and reduced the deviation between the fold compared to the IMU.

It was discovered that hand coordinates alone enabled to achieve the highest accuracy to detect all the activities. However, this dataset was biased to perform well on this data as each position was associated with a certain activity. However, with current technology it requires 4 IMUs on torso, upper and lower arms, and hand. It is not convenient to use this set up on a large number of people.

For the best performance on this dataset the hand coordinates with 5 features should be used to achieve highest accuracy of 80.7% in detecting scratches. On the other hand, if comfortable system is a priority and accuracy can be sacrificed, then 70% accuracy can be achieved with a single device using IMU and EPS.

8 ACKNOWLEDGEMENTS

This work was partly funded by EPSRC grant EP/S513921/1.

REFERENCES

- [1] [n.d.]. Seconds.
- [2] Diana Bautista, Sarah Wilson, and Mark Hoon. 2014. Why we scratch an itch: The molecules, cells and circuits of itch. *Nature neuroscience* 17 (02 2014), 175–82.
- [3] Akram Bayat, Marc Pomplun, and Duc A. Tran. 2014. A Study on Human Activity Recognition Using Accelerometer Data from Smartphones. *Procedia Computer Science* 34 (2014), 450 – 457. <http://www.sciencedirect.com/science/article/pii/S1877050914008643> The 9th International Conference on Future Networks and Communications (FNC'14)/The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC'14)/Affiliated Workshops.
- [4] F. Bonnin-Pascual and A. Ortiz. 2019. An UWB-based System for Localization inside Merchant Vessels. In *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. 1559–1562.
- [5] Thomas M. Cover and Joy A. Thomas. 2006. *Elements of Information Theory* (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, USA.

- [6] Toshiya Ebata, Hiroshi Aizawa, and Ryoichi Kamide. 1996. An Infrared Video Camera System to Observe Nocturnal Scratching in Atopic Dermatitis Patients. *The Journal of Dermatology* 23, 3 (1996), 153–155. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1346-8138.1996.tb03990.x>
- [7] T Ebata, S Iwasaki, R Kamide, and M Niimura. 2001. Use of a wrist activity monitor for the measurement of nocturnal scratching in patients with atopic dermatitis. *The British journal of dermatology* 144 (02 2001), 305–9.
- [8] Kaoru Endo, Hozumi Sano, Takayuki Fukuzumi, Jun Adachi, and Toshiyuki Aoki. 1999. Objective scratch monitor evaluation of the effect of an antihistamine on nocturnal scratching in atopic dermatitis. *Journal of Dermatological Science* 22, 1 (1999), 54 – 61. <http://www.sciencedirect.com/science/article/pii/S0923181199000481>
- [9] Jongin Lee, Dae-ki Cho, Seokwoo Song, SeungHo Kim, Eunji Im, and John Kim. 2015. Mobile System Design for Scratch Recognition. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1567–1572. <https://doi.org/10.1145/2702613.2732820>
- [10] A. Moreau, P. Anderer, M. Ross, A. Cerny, T. H. Almazan, and B. Peterson. 2018. Detection of Nocturnal Scratching Movements in Patients with Atopic Dermatitis Using Accelerometers and Recurrent Neural Networks. *IEEE Journal of Biomedical and Health Informatics* 22, 4 (July 2018), 1011–1018.
- [11] Yuichi Noro, Youichi Omoto, Koji Umeda, Futa Tanaka, Yousuke Shiratsuka, Tomomi Yamada, Kenichi Isoda, Kimiko Matsubara, Keiichi Yamanaka, Esteban C. Gabazza, Masakatsu Nishikawa, and Hitoshi Mizutani. 2014. Novel acoustic evaluation system for scratching behavior in itching dermatitis: Rapid and accurate analysis for nocturnal scratching of atopic dermatitis patients. *The Journal of Dermatology* 41, 3 (2014), 233–238. [arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/1346-8138.12405](https://onlinelibrary.wiley.com/doi/pdf/10.1111/1346-8138.12405) <https://onlinelibrary.wiley.com/doi/abs/10.1111/1346-8138.12405>
- [12] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* 16, 1 (2016). <https://www.mdpi.com/1424-8220/16/1/115>
- [13] J. Petersen, D. Austin, R. Sack, and T. L. Hayes. 2013. Actigraphy-Based Scratch Detection Using Logistic Regression. *IEEE Journal of Biomedical and Health Informatics* 17, 2 (March 2013), 277–283.
- [14] A. Pouryazdan, R. J. Prance, H. Prance, and D. Roggen. 2016. Wearable Electric Potential Sensing: A New Modality Sensing Hair Touch and Restless Leg Movement. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 846–850. <http://doi.acm.org/10.1145/2968219.2968286>
- [15] Daniel Roggen, Arash Pouryazdan, and Mathias Ciliberto. 2018. Poster: BlueSense - Designing an Extensible Platform for Wearable Motion Sensing, Sensor Research and IoT Applications. In *Proceedings of the 2018 International Conference on Embedded Wireless Systems and Networks (EWSN ’18)*. Junction Publishing, USA, 177–178. <http://dl.acm.org/citation.cfm?id=3234847.3234874>
- [16] P. M. Scholl and K. van Laerhoven. 2012. A Feasibility Study of Wrist-Worn Accelerometer Based Detection of Smoking Habits. In *2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*. 886–891. <https://doi.org/10.1109/IMIS.2012.96>
- [17] A. Torralba and A. Oliva. 2002. Depth Estimation from Image Structure. *IEEE Transactions on Pattern Analysis Machine Intelligence* 27, 09 (sep 2002), 1226–1238.
- [18] R. Twycross, M.W. Greaves, H. Handwerker, E.A. Jones, S.E. Libretto, J.C. Szepletowski, and Z. Zylicz. 2003. Itch: scratching more than the surface. *QJM: An International Journal of Medicine* 96, 1 (01 2003), 7–26. [arXiv:http://oup.prod.sis.lan/qjmed/article-pdf/96/1/7/4375153/960007.pdf](http://oup.prod.sis.lan/qjmed/article-pdf/96/1/7/4375153/960007.pdf) <https://doi.org/10.1093/qjmed/hcg002>